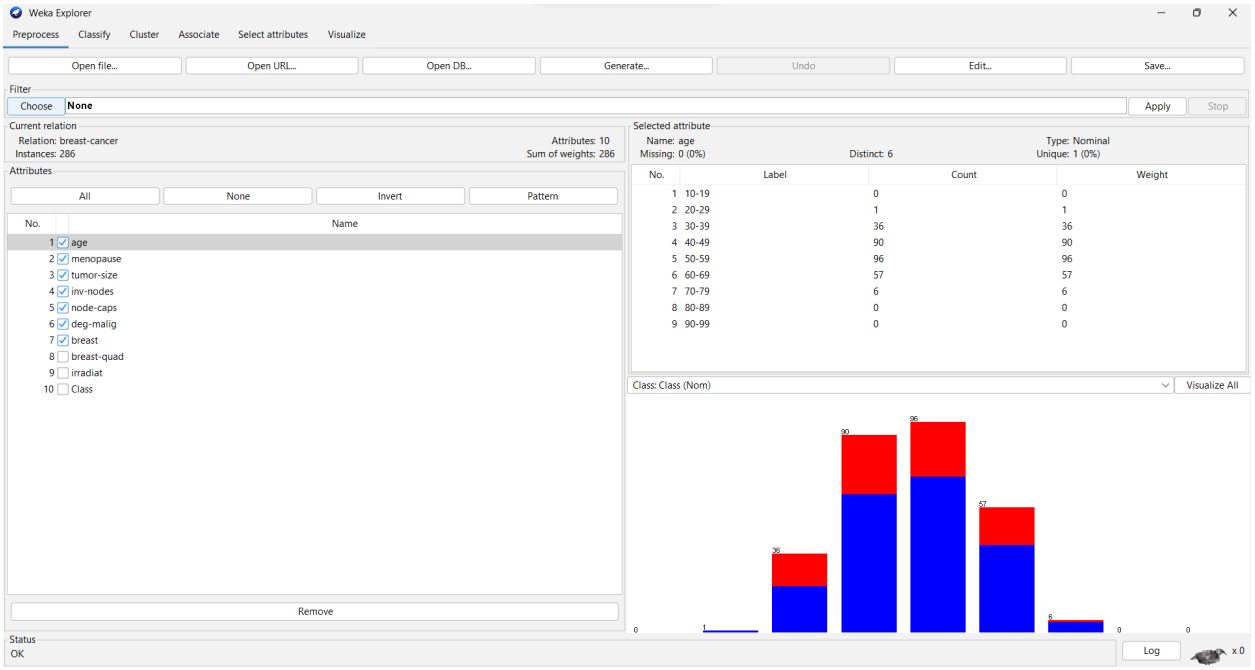
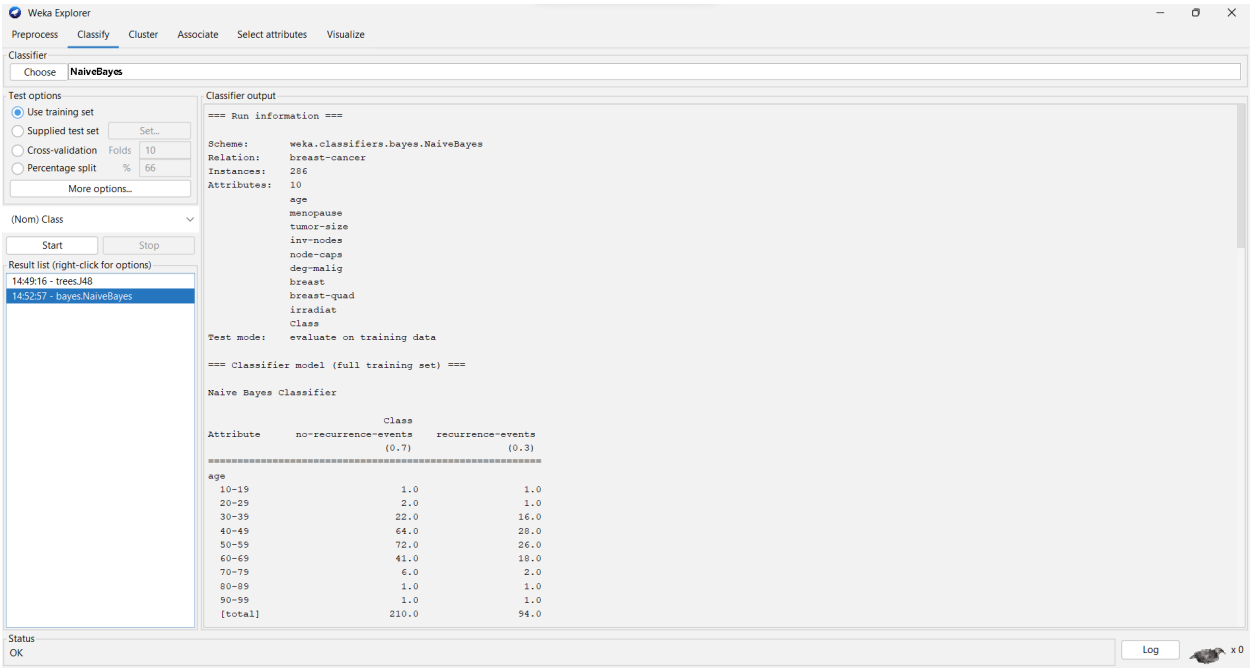
Name: Swapnil Santosh Sawant
Roll no: 47
Division: TE4 Batch : D
Subject: Data Warehousing and Mining

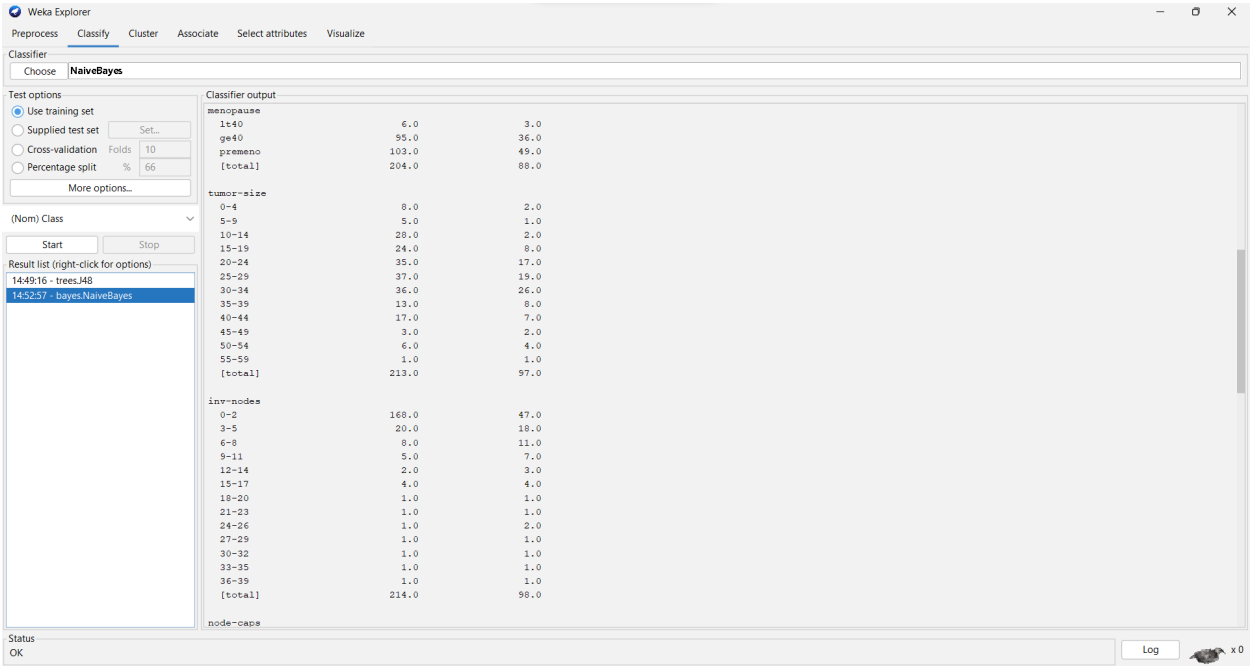
Experiment-6

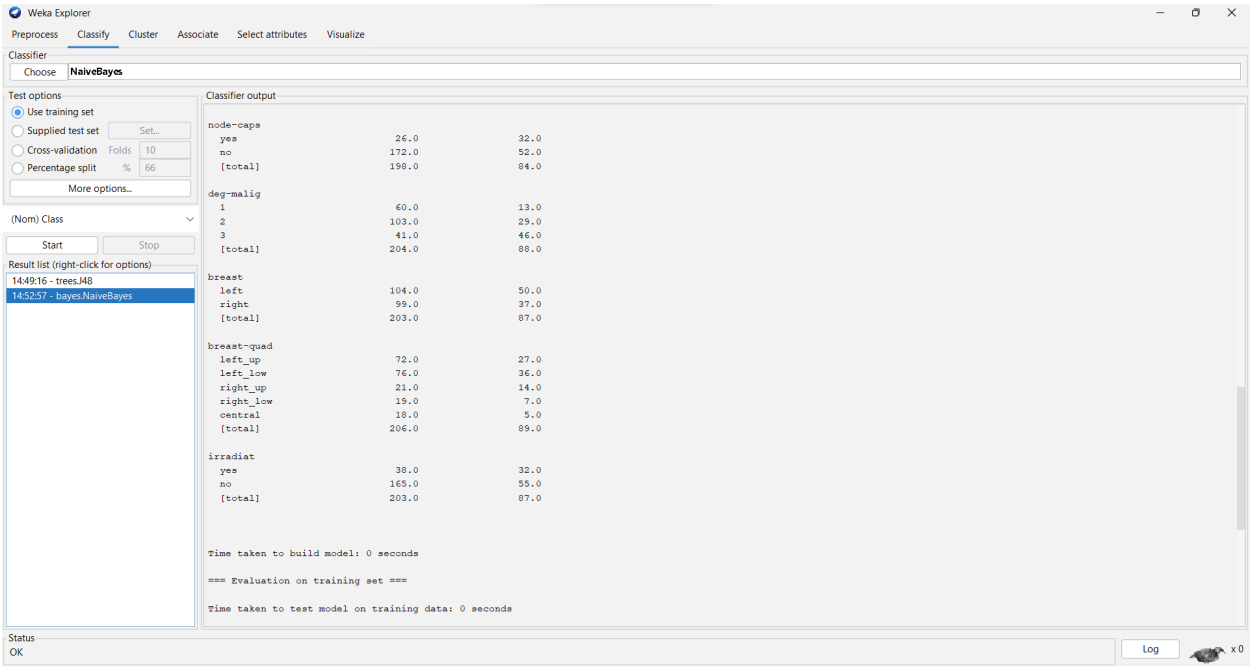
i) Weka Classification using Naive Bayesian

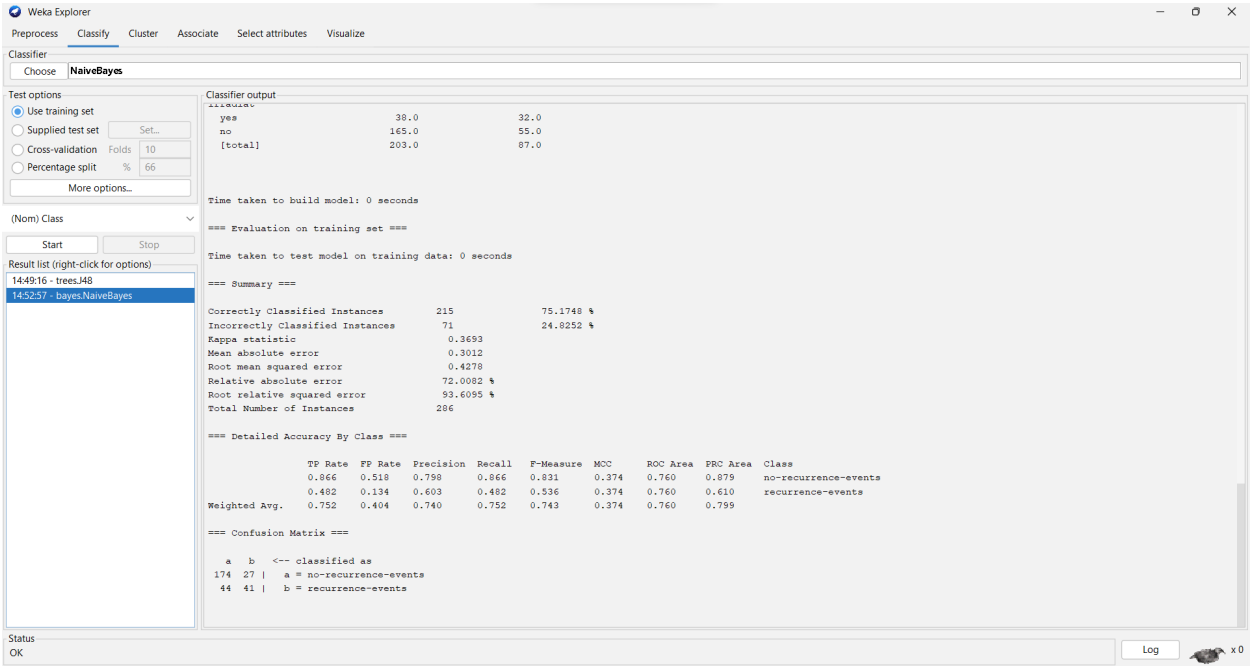
Explorer:

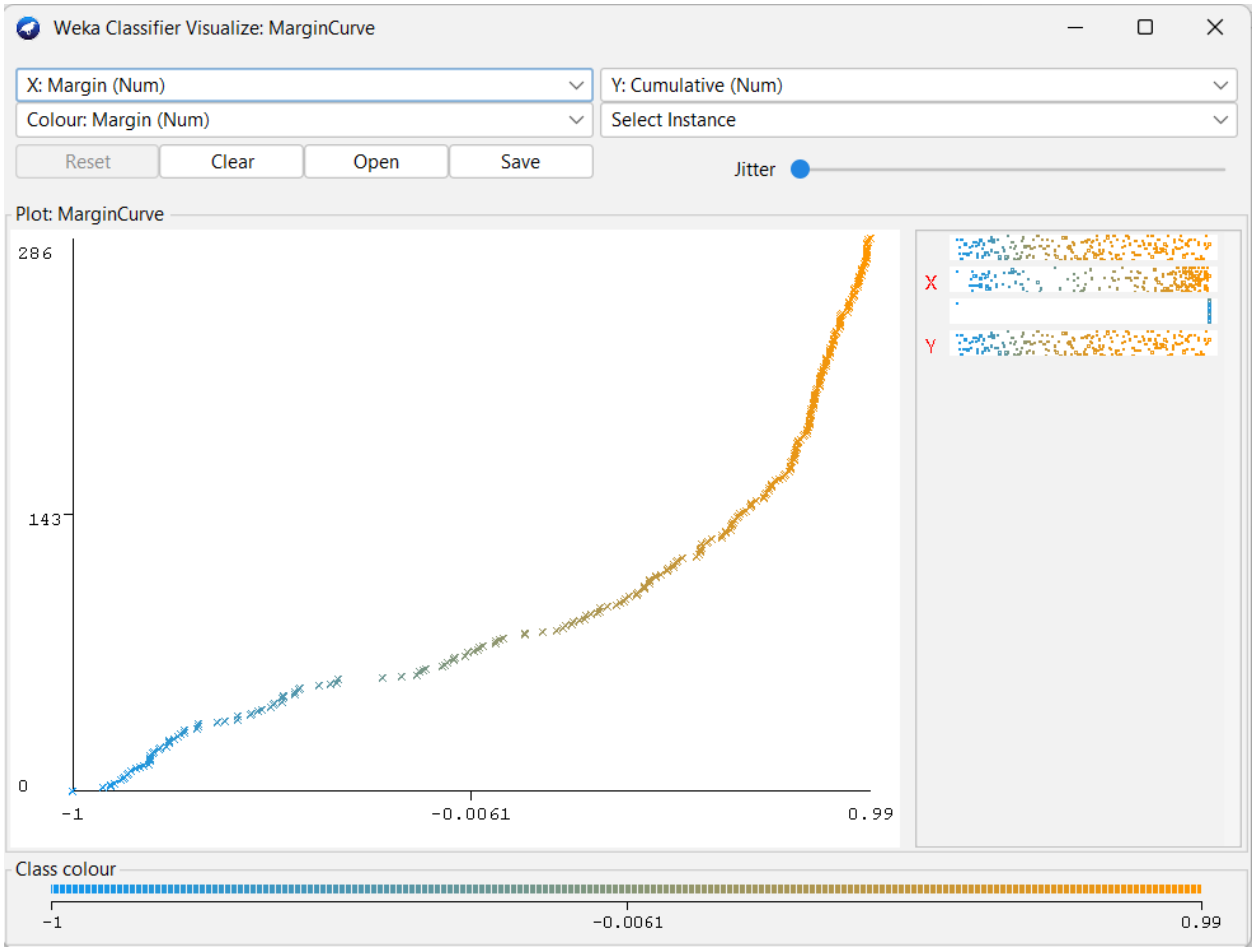



Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
NaiveBayes
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) Class
Start
Stop
Result list (right-click for options)
14:49:16 - trees.J48
14:52:57 - bayes.NaiveBayes
Classifier output
=== Run information ===
Scheme: weka.classifiers.bayes.NaiveBayes
Relation: breast-cancer
Instances: 286
Attributes: 10
age
menopause
tumor-size
inv-nodes
node-caps
deg-malig
breast
breast-quad
irradiat
Class
Test mode: evaluate on training data
=== Classifier model (full training set) ===
Naive Bayes Classifier
Class
Attribute no-recurrence-events recurrence-events
(0.7) (0.3)
age
10-19 1.0 1.0
20-29 2.0 1.0
30-39 22.0 16.0
40-49 64.0 28.0
50-59 72.0 26.0
60-69 41.0 18.0
70-79 6.0 2.0
80-89 1.0 1.0
90-99 1.0 1.0
[total] 210.0 94.0
Status
OK
Log
x 0

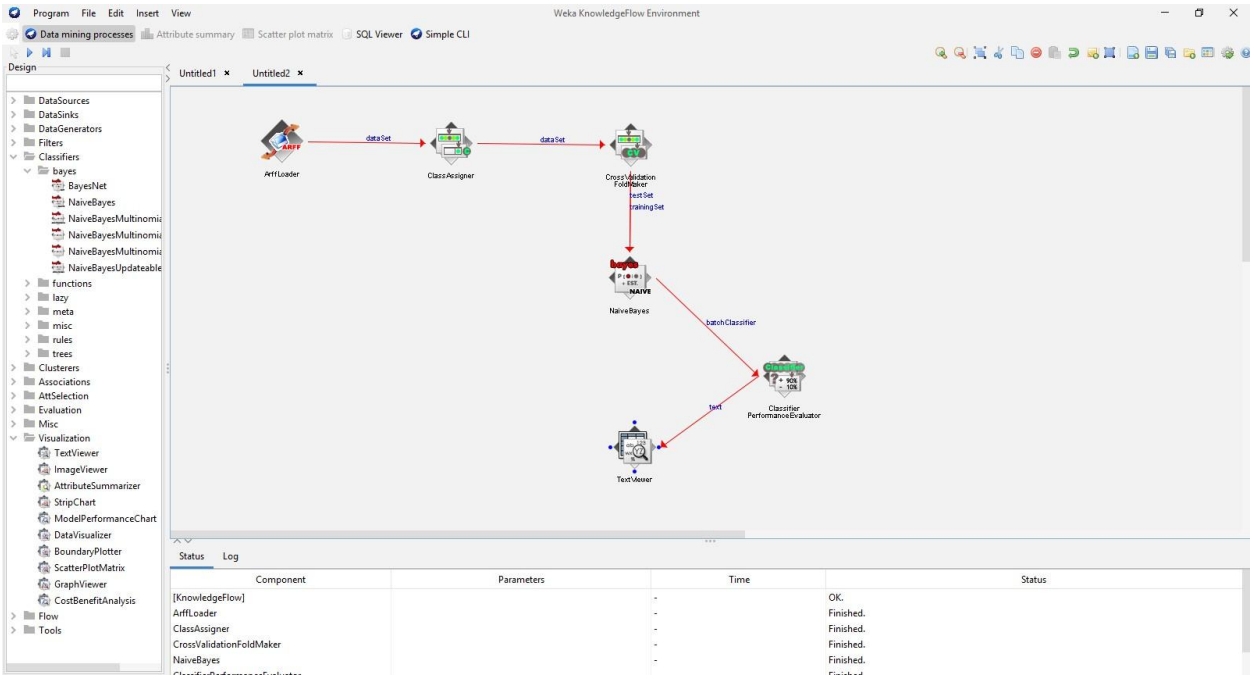

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
NaiveBayes
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) Class
Start
Stop
Result list (right-click for options)
14:49:16 - trees.J48
14:52:57 - bayes.NaiveBayes
Classifier output
menopause
lt40 6.0 3.0
ge40 95.0 36.0
premeno 103.0 49.0
[total] 204.0 88.0
tumor-size
0-4 8.0 2.0
5-9 5.0 1.0
10-14 28.0 2.0
15-19 24.0 8.0
20-24 35.0 17.0
25-29 37.0 19.0
30-34 36.0 26.0
35-39 13.0 8.0
40-44 17.0 7.0
45-49 3.0 2.0
50-54 6.0 4.0
55-59 1.0 1.0
[total] 213.0 97.0
inv-nodes
0-2 168.0 47.0
3-5 20.0 18.0
6-8 8.0 11.0
9-11 5.0 7.0
12-14 2.0 3.0
15-17 4.0 4.0
18-20 1.0 1.0
21-23 1.0 1.0
24-26 1.0 2.0
27-29 1.0 1.0
30-32 1.0 1.0
33-35 1.0 1.0
36-39 1.0 1.0
[total] 214.0 98.0
node-caps
Status
OK
Log
x 0

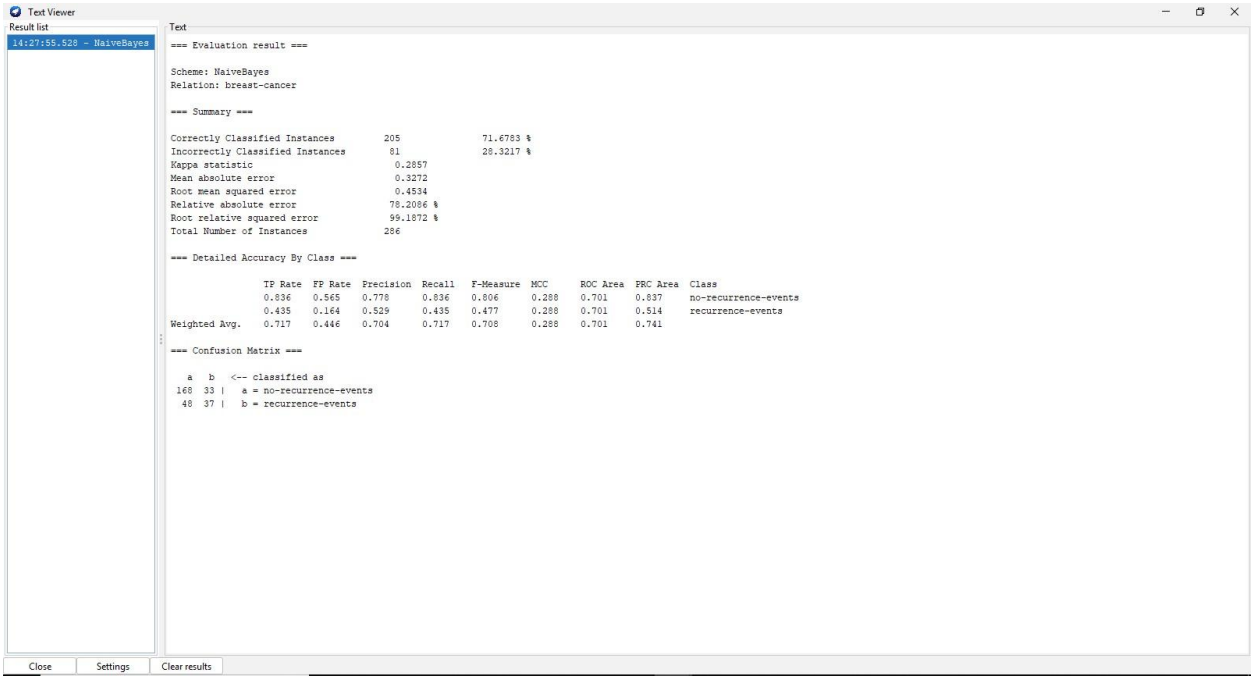

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
NaiveBayes
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) Class
Start
Stop
Result list (right-click for options)
14:49:16 - trees.J48
14:52:57 - bayes.NaiveBayes
Classifier output
node-caps
yes 26.0 32.0
no 172.0 52.0
[total] 198.0 84.0
deg-malig
1 60.0 13.0
2 103.0 29.0
3 41.0 46.0
[total] 204.0 88.0
breast
left 104.0 50.0
right 99.0 37.0
[total] 203.0 87.0
breast-quad
left_up 72.0 27.0
left_low 76.0 36.0
right_up 21.0 14.0
right_low 19.0 7.0
central 18.0 5.0
[total] 206.0 89.0
irradiat
yes 38.0 32.0
no 165.0 55.0
[total] 203.0 87.0
Time taken to build model: 0 seconds
=== Evaluation on training set ===
Time taken to test model on training data: 0 seconds
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
NaiveBayes
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) Class
Start
Stop
Result list (right-click for options)
14:49:16 - trees.J48
14:52:57 - bayes.NaiveBayes
Classifier output
irradiat
yes 38.0 32.0
no 165.0 55.0
[total] 203.0 87.0
Time taken to build model: 0 seconds
=== Evaluation on training set ===
Time taken to test model on training data: 0 seconds
=== Summary ===
Correctly Classified Instances 215 75.1748 %
Incorrectly Classified Instances 71 24.8252 %
Kappa statistic 0.3693
Mean absolute error 0.3012
Root mean squared error 0.4278
Relative absolute error 72.0082 %
Root relative squared error 93.6095 %
Total Number of Instances 286
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
0.866 0.518 0.798 0.866 0.831 0.374 0.760 0.879 no-recurrence-events
0.482 0.134 0.603 0.482 0.536 0.374 0.760 0.610 recurrence-events
Weighted Avg. 0.752 0.404 0.740 0.752 0.743 0.374 0.760 0.799
=== Confusion Matrix ===
a b <-- classified as
174 27 | a = no-recurrence-events
44 41 | b = recurrence-events
Status
OK
Log
x 0


Weka Classifier Visualize: MarginCurve
X: Margin (Num)
Y: Cumulative (Num)
Colour: Margin (Num)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot: MarginCurve
286
143
0
-1
-0.0061
0.99
X
Y
Class colour
-1
-0.0061
0.99

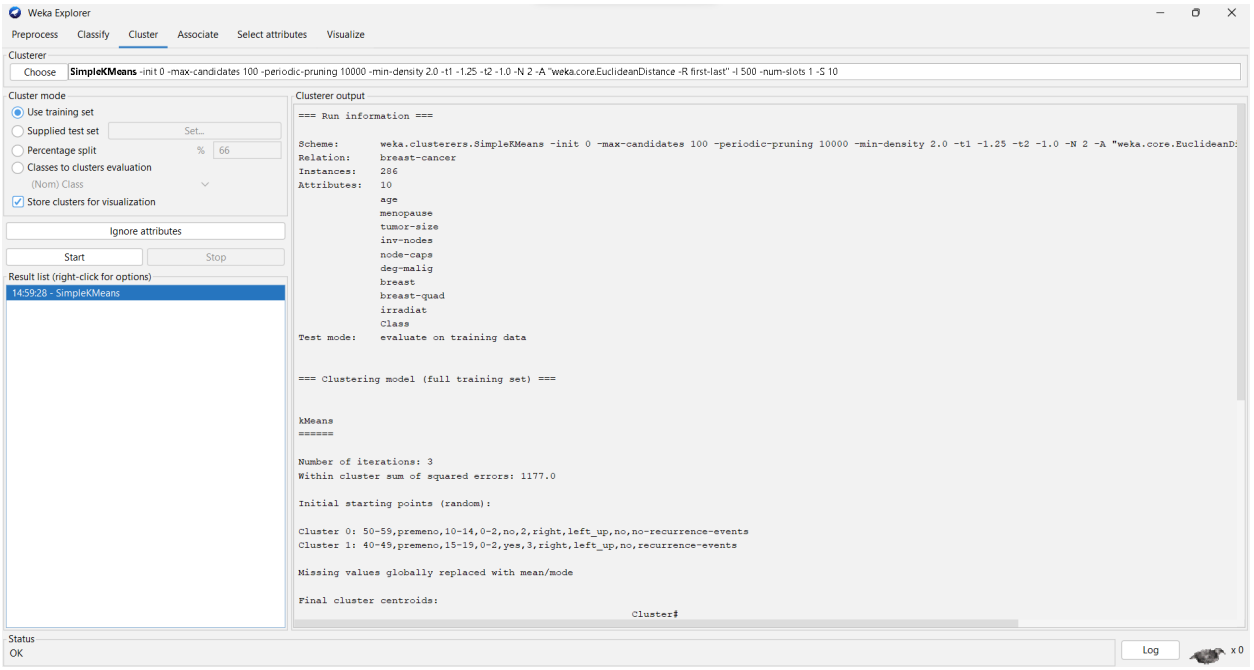
Knowledge Base:

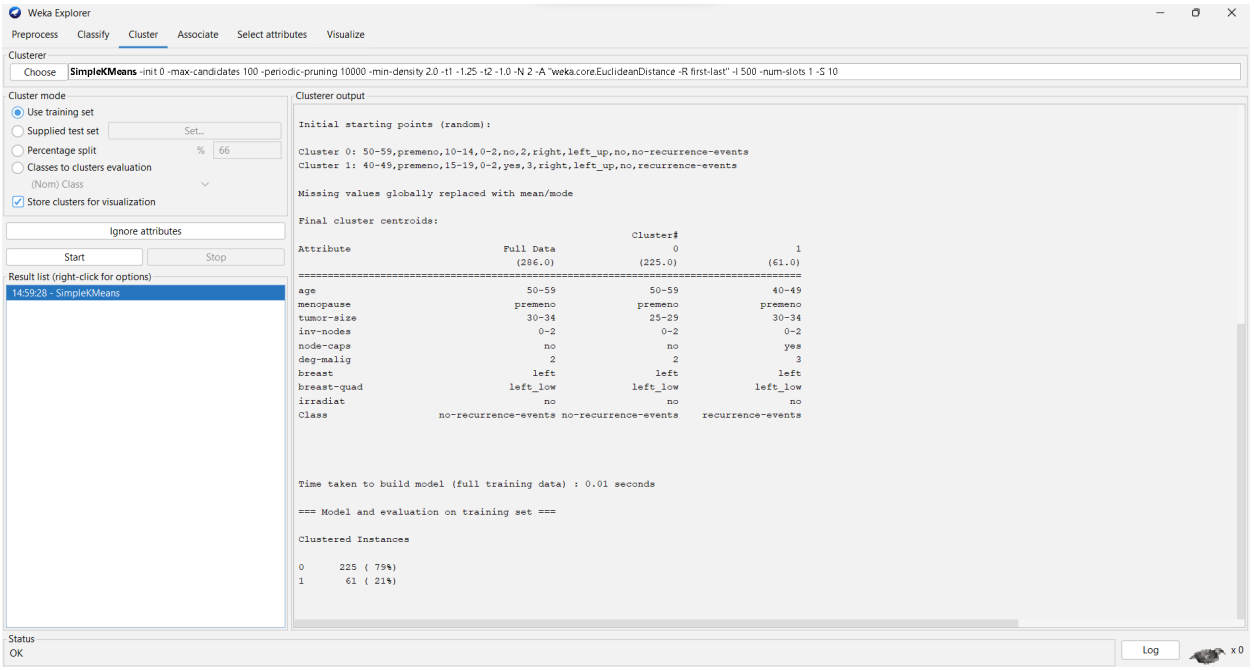


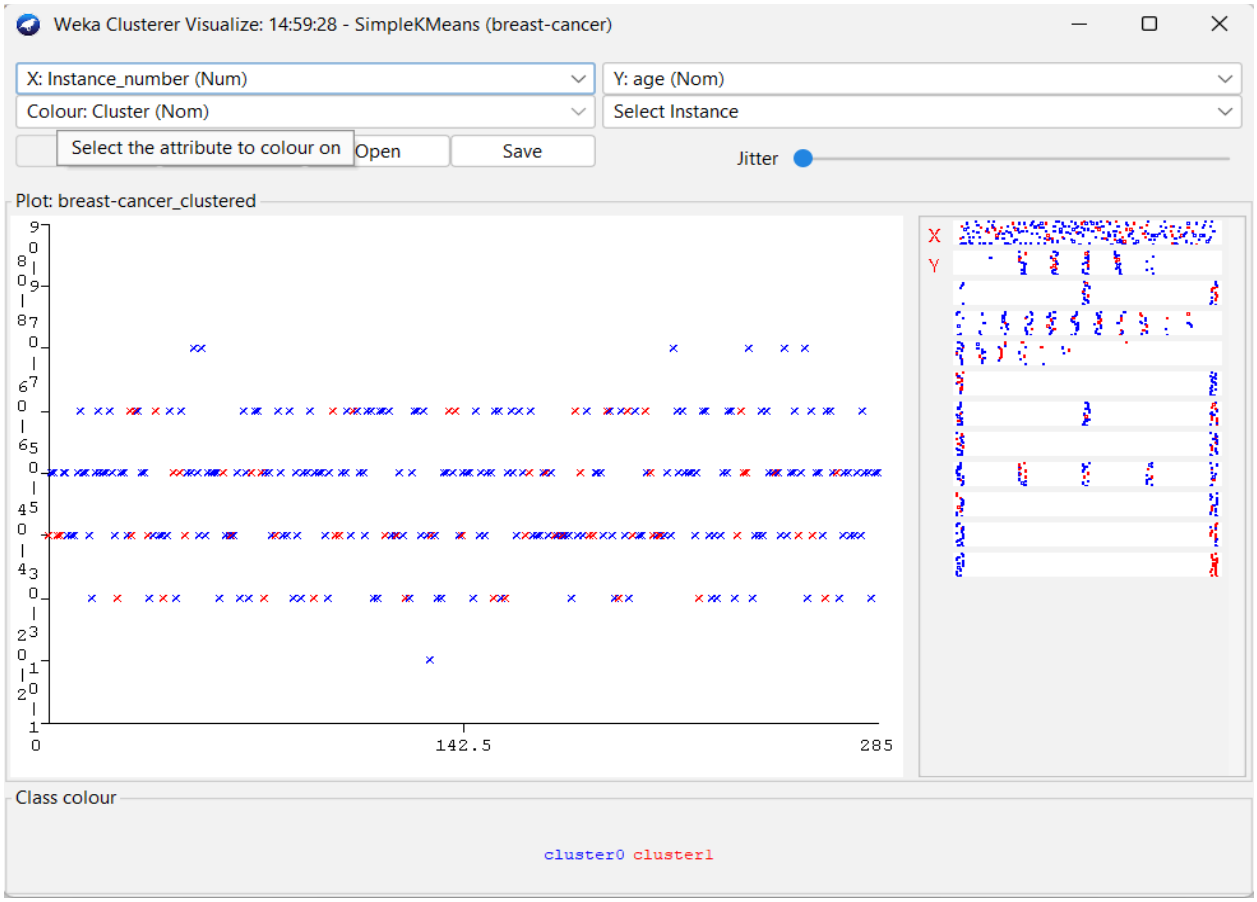


ii) Weka Clustering using K-means:

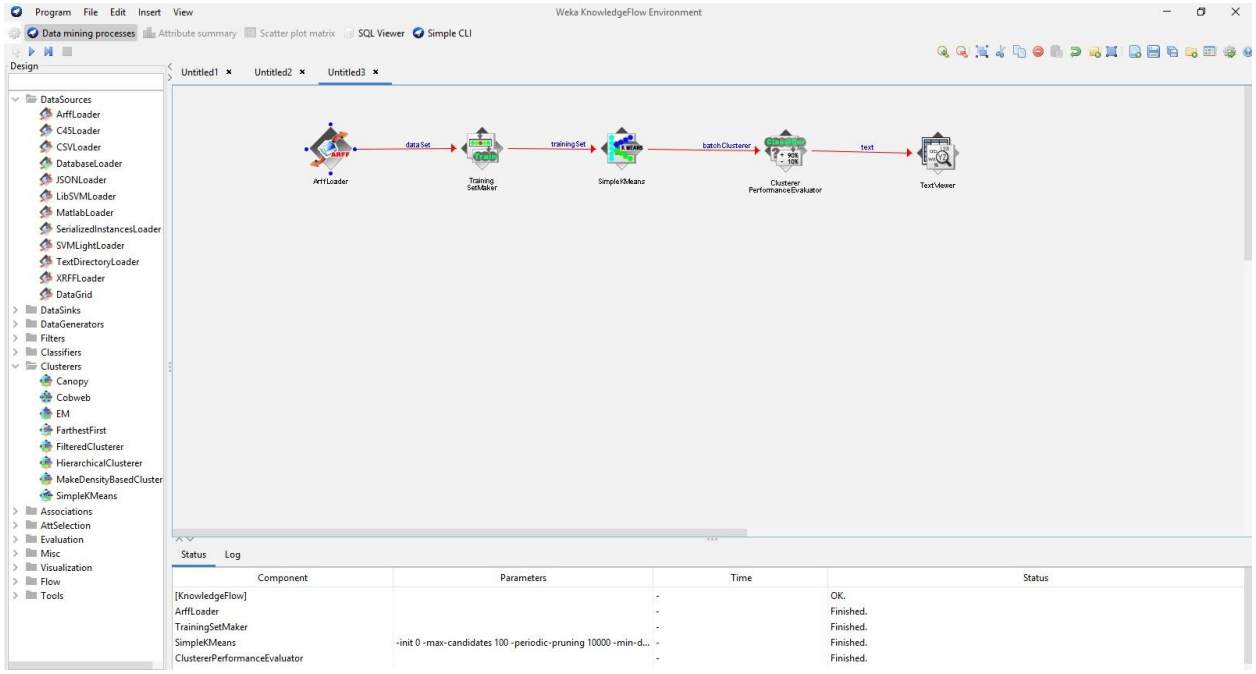
Explorer:

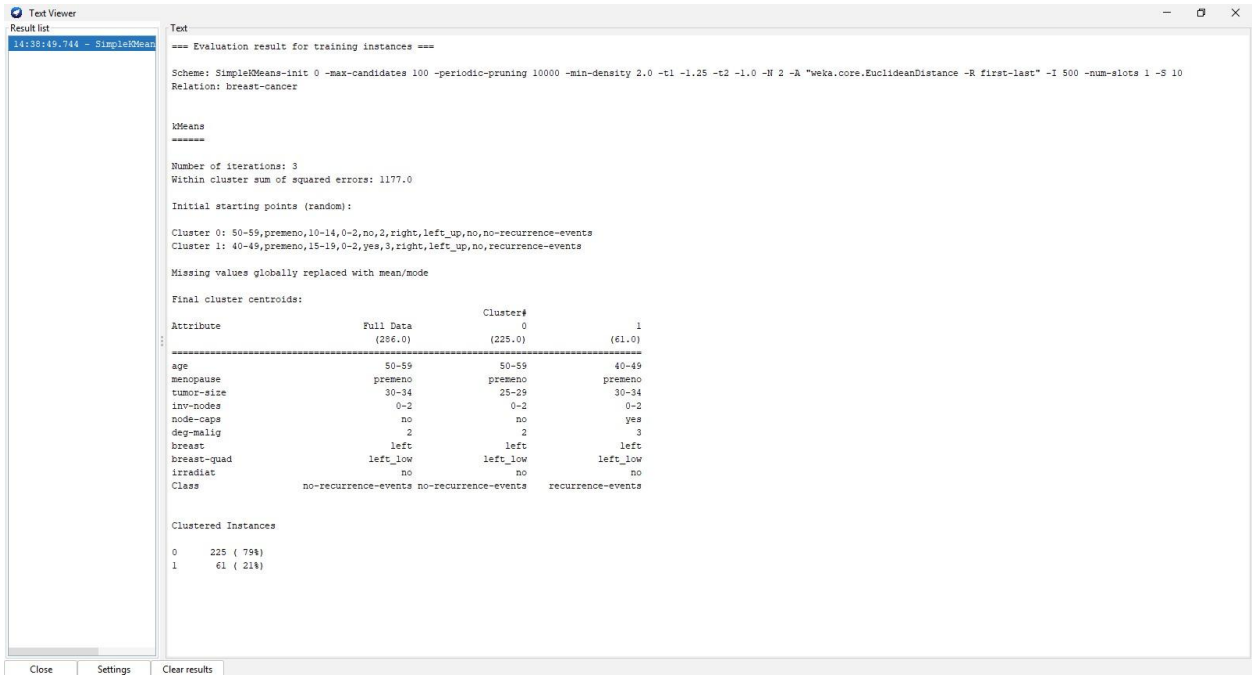






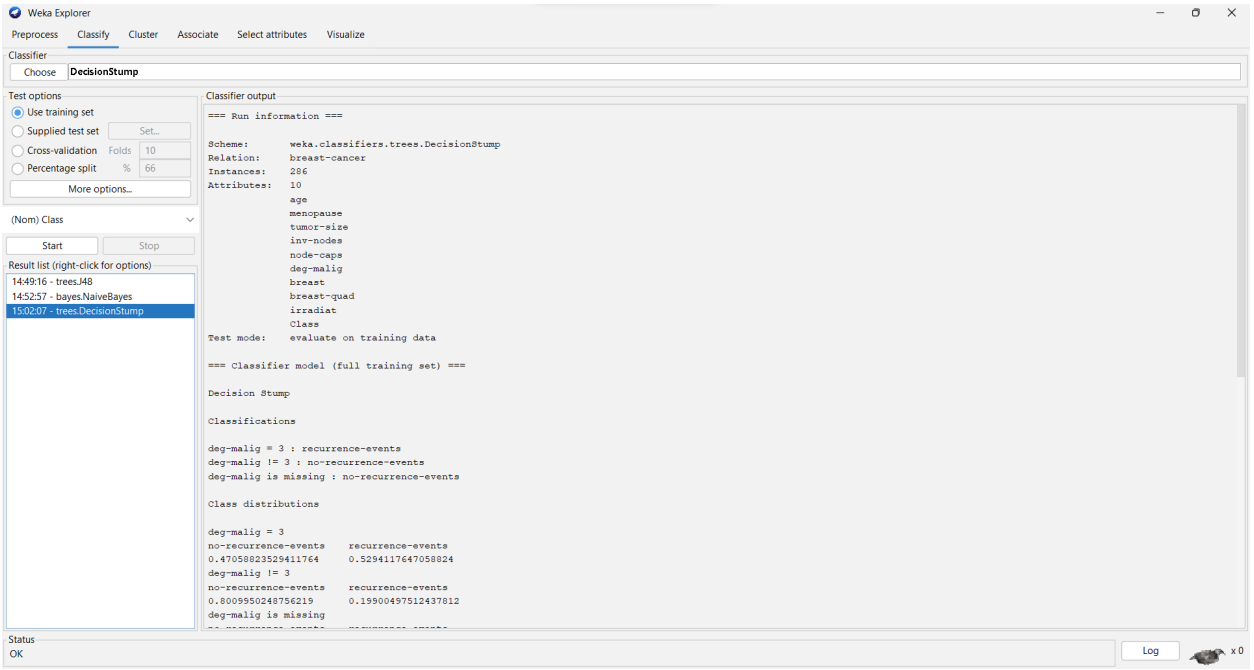
Knowledge base:

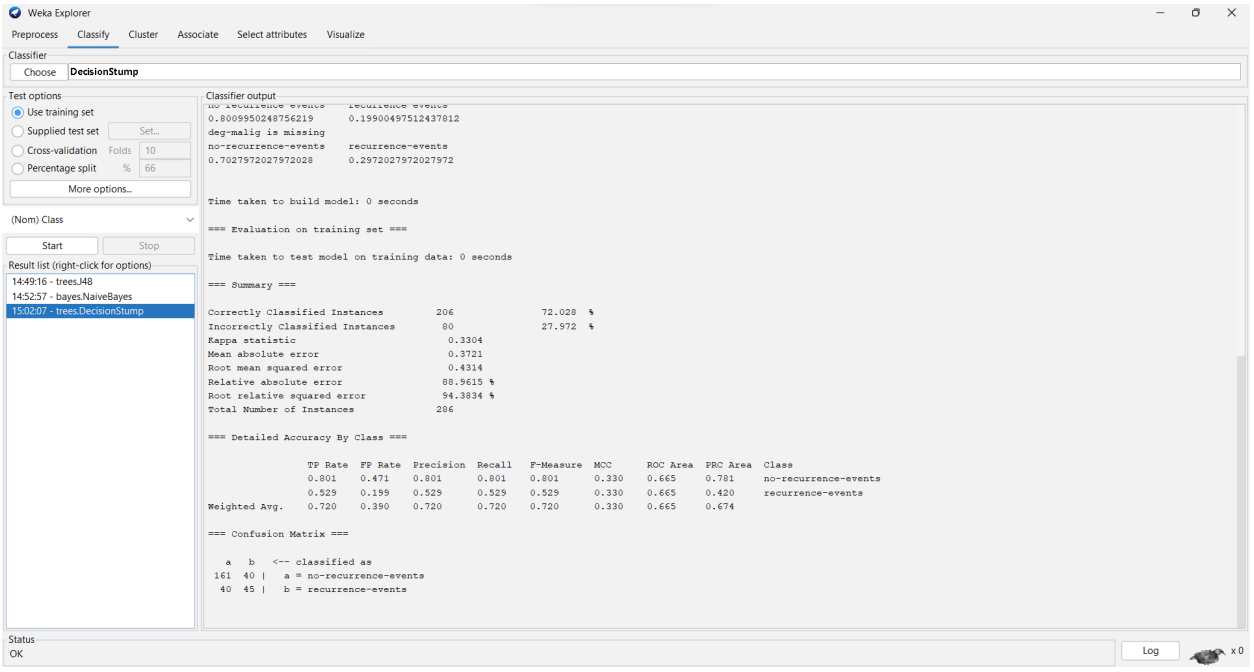



Text Viewer
Result list
14:38:49.744 - SimpleKMean
Text
=== Evaluation result for training instances ===
Scheme: SimpleKMeans-init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10
Relation: breast-cancer
kMeans
======
Number of iterations: 3
Within cluster sum of squared errors: 1177.0
Initial starting points (random):
Cluster 0: 50-59,premeno,10-14,0-2,no,2,right,left_up,no,no-recurrence-events
Cluster 1: 40-49,premeno,15-19,0-2,yes,3,right,left_up,no,recurrence-events
Missing values globally replaced with mean/mode
Final cluster centroids:
Cluster#
Attribute Full Data 0 1
(286.0) (225.0) (61.0)
age 50-59 50-59 40-49
menopause premeno premeno premeno
tumor-size 30-34 25-29 30-34
inv-nodes 0-2 0-2 0-2
node-caps no no yes
deg-malig 2 2 3
breast left left left
breast-quad left_low left_low left_low
irradiat no no no
Class no-recurrence-events no-recurrence-events recurrence-events
Clustered Instances
0 225 (79%)
1 61 (21%)
Close
Settings
Clear results

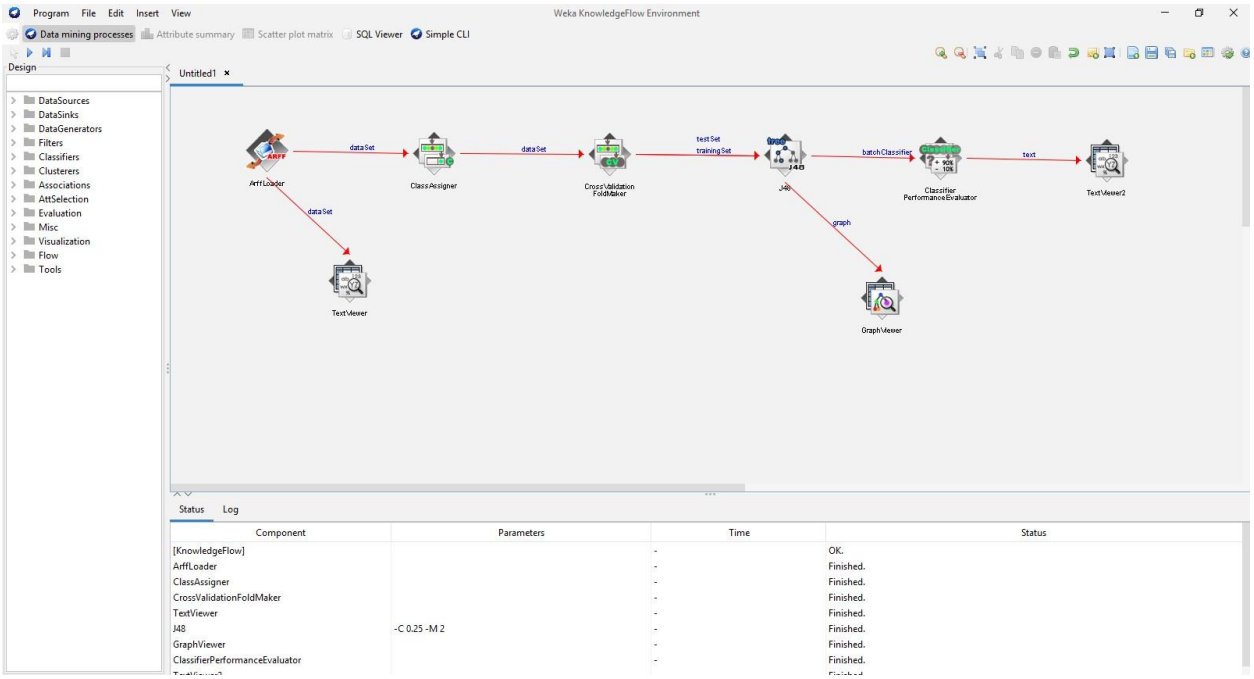
iii)Weka Classification using Decision tree:

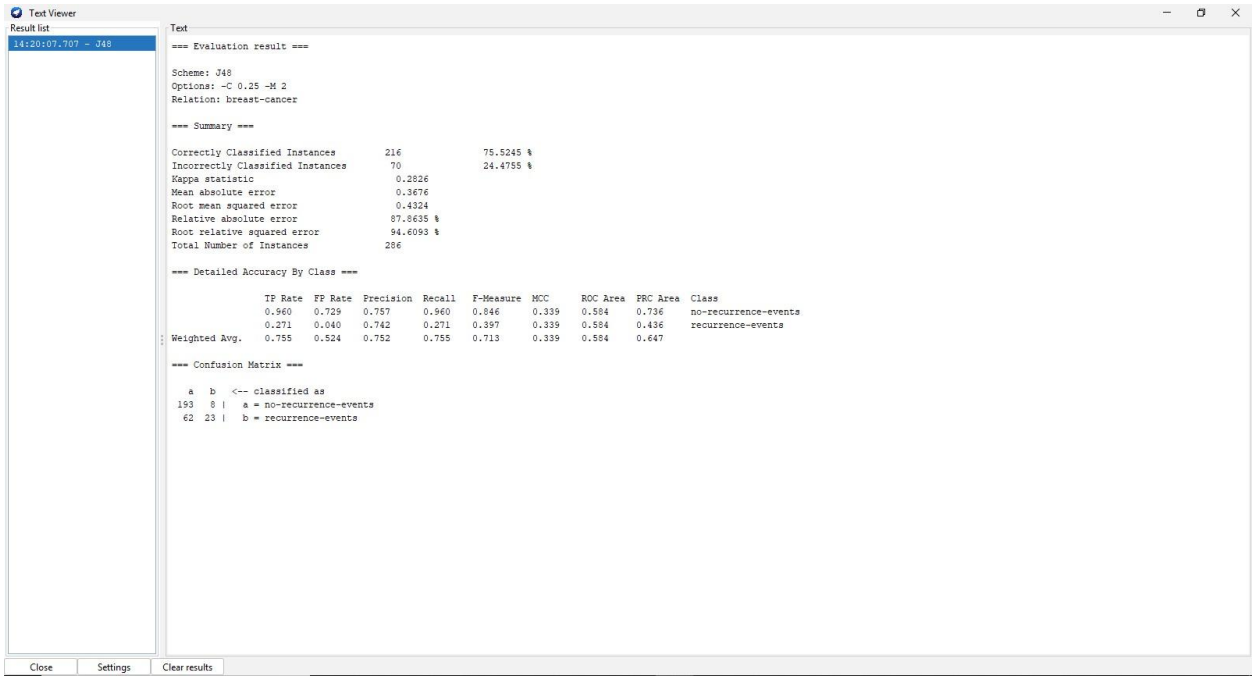
Explorer:

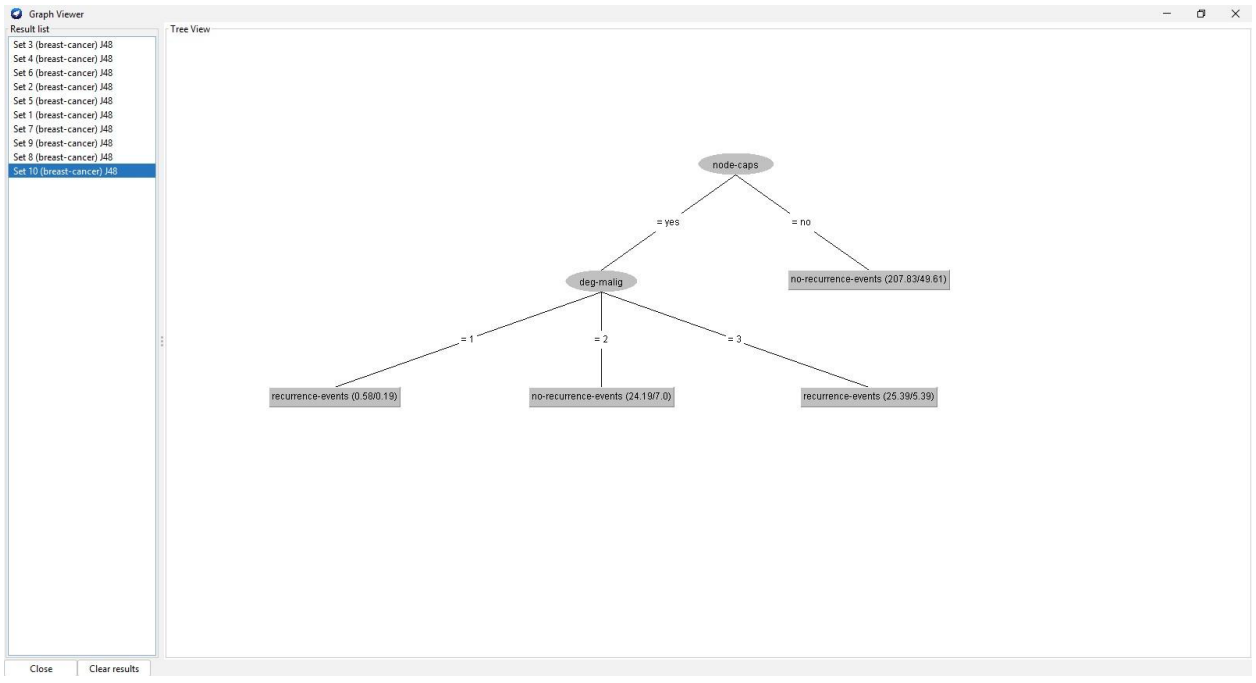




Knowledge base:

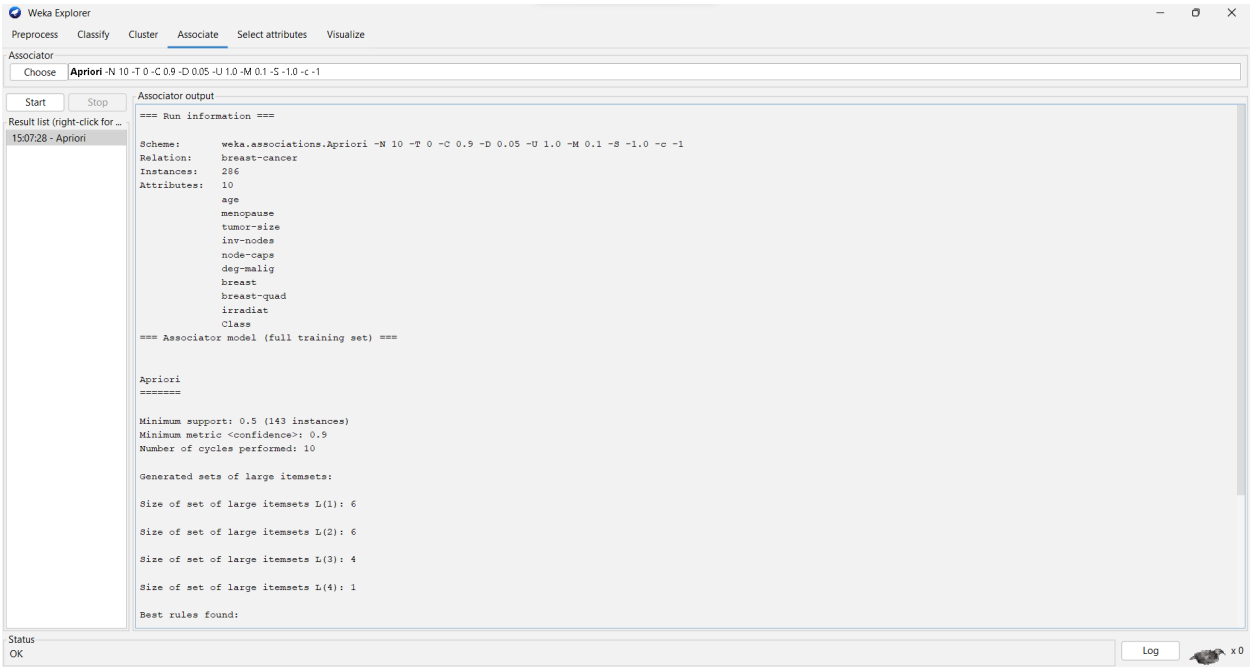


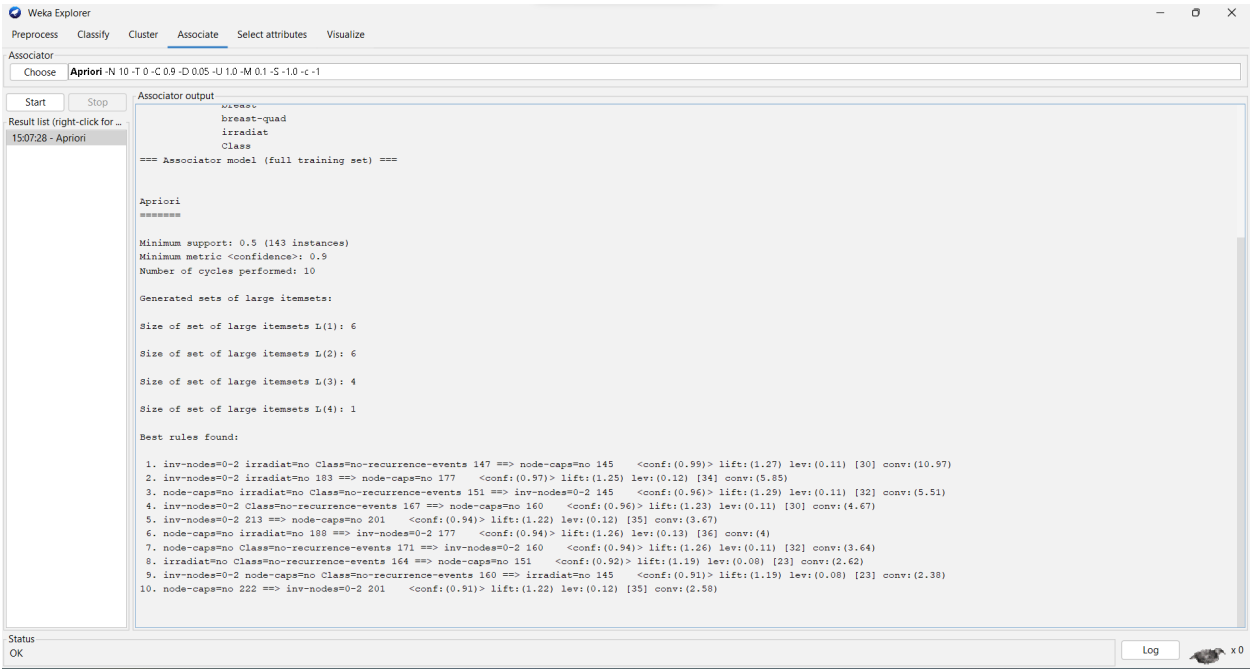

Text Viewer
Result list
14:20:07.707 - J48
Text
=== Evaluation result ===
Scheme: J48
Options: -C 0.25 -M 2
Relation: breast-cancer
=== Summary ===
Correctly Classified Instances 216 75.5245 %
Incorrectly Classified Instances 70 24.4755 %
Kappa statistic 0.2826
Mean absolute error 0.3676
Root mean squared error 0.4324
Relative absolute error 87.8635 %
Root relative squared error 94.6093 %
Total Number of Instances 286
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
0.960 0.729 0.757 0.960 0.846 0.339 0.584 0.736 no-recurrence-events
0.271 0.040 0.742 0.271 0.397 0.339 0.584 0.436 recurrence-events
Weighted Avg. 0.755 0.524 0.752 0.755 0.713 0.339 0.584 0.647
=== Confusion Matrix ===
a b <-- classified as
193 8 | a = no-recurrence-events
62 23 | b = recurrence-events
Close
Settings
Clear results


Graph Viewer
Result list
Set 3 (breast-cancer) J48
Set 4 (breast-cancer) J48
Set 6 (breast-cancer) J48
Set 2 (breast-cancer) J48
Set 5 (breast-cancer) J48
Set 1 (breast-cancer) J48
Set 7 (breast-cancer) J48
Set 9 (breast-cancer) J48
Set 8 (breast-cancer) J48
Set 10 (breast-cancer) J48
Tree View
node-caps
= yes
= no
deg-malig
no-recurrence-events (207.83/49.61)
= 1
= 2
= 3
recurrence-events (0.58/0.19)
no-recurrence-events (24.19/7.0)
recurrence-events (25.39/5.39)
Close
Clear results

iv) Weka Association using Apriori :

Explorer:





Knowledge Base: